Medical Expenditure Panel Survey

HOUSEHOLD COMPONENT
LINKING MEPS DATA

Types of Linking to be Covered

- **Linking person-level file to external data sources**
  - **National Health Interview Survey PUF files**
  - **Area Resource File and other secondary data**
- **Linking person-level file to other MEPS files**
  - **Events files**
  - **Conditions file**
  - **Conditions file to the events file**
  - **Jobs file**

Person-level files include the Point-in-Time file, the Full Year Population Characteristics file, and the Full Year Consolidated Data File.

Universal MEPS Link

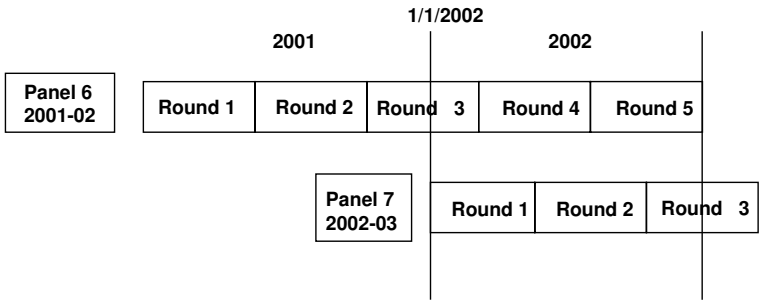
- DUPERSID uniquely identifies each MEPS sample person and is created from DUID + PID.
- Remains the same for survey duration.
- Is the link between all MEPS data files.

Linking to the NHIS

- The MEPS-HC uses the NHIS respondents as its sampling frame.
- Each year a new MEPS-HC panel is established drawing from the previous year's NHIS sample .
- MEPS data collection utilizes an overlapping panel design.

Linking to the NHIS

Panel 6 Year 2 & Panel 7 Year 1



For full calendar year 2002 estimates, Rounds 1, 2, and 3 of Panel 7 (which uses the 2001 NHIS as its sampling frame) are combined with Rounds 3, 4, and 5 of Panel 6 (which uses the 2000 NHIS as its sampling frame).

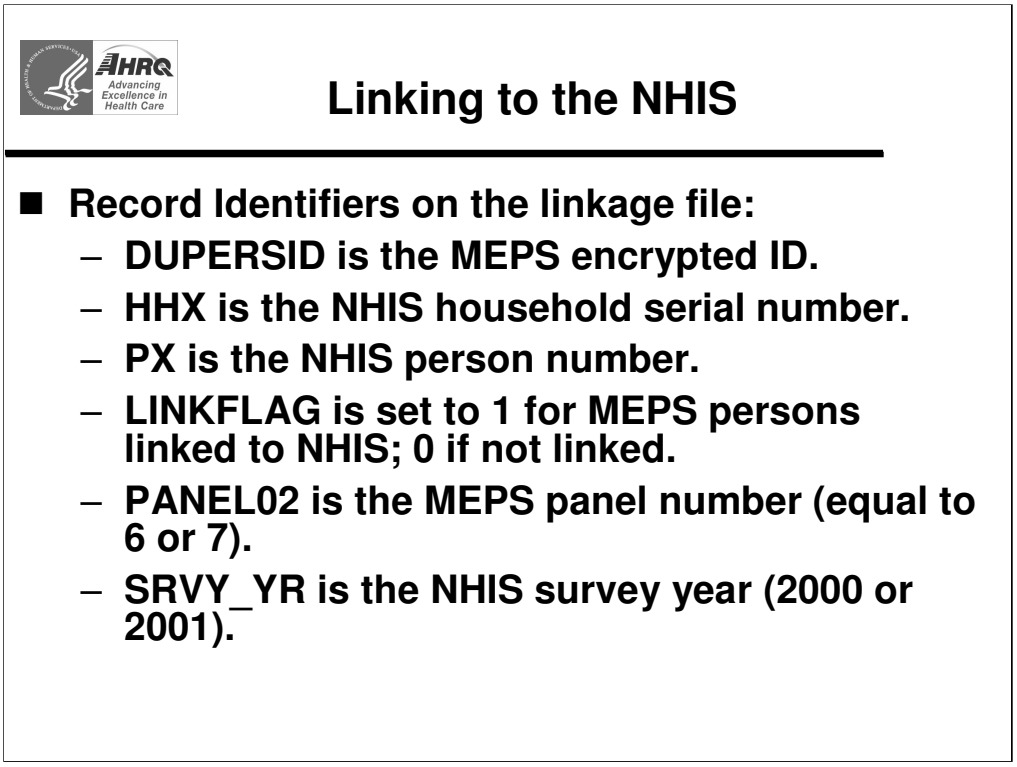
Linking to the NHIS

- Each full-year file contains sample from two different NHIS years.
- Every MEPS file has a panel indicator variable to identify the year it was first fielded.
- NHIS link file required to link MEPS and NHIS PUFs.
- Request link file at mepspd@ahrq.gov.

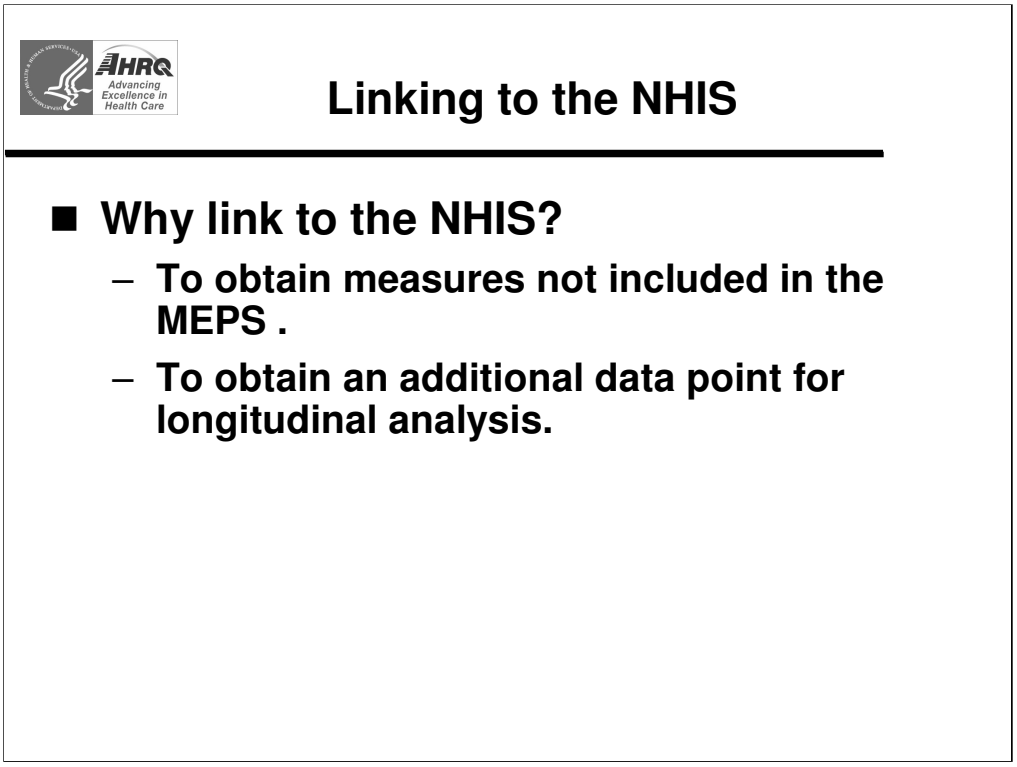
Linking to the NHIS

- **Link files are only cross-walks and contain no actual data.**
- **NCHS PUF files can be downloaded from the CDC Web site: http://www.cdc.gov/nchs/nhis.htm**
- **Only variables from the Family Section of the NHIS should be used.**

When selecting variables from the NHIS to merge onto the MEPS, only variables from the Family section of the NHIS should be considered, since these variables will have been asked of all persons. Supplements only collect data from a sampled adult or child.



MEPS started in 1996 and panels are numbered sequentially according to start year; i.e., Panel 1 started in 1996, Panel 2 started in 1997, etc.



For example, to track health status changes over time for a particular condition, you may want to look at changes in health status over three years.

Linked MEPS-NHIS Estimation Issues

- Estimates are based on MEPS population and utilize MEPS weights.
- There are persons in the MEPS that are not in the NHIS:
  - Children born after the NHIS interview
  - Persons that married into the family after the NHIS interview
  - Persons released from an institution after the NHIS interview

Linked MEPS-NHIS Estimation Issues

- There are also persons in the NHIS that are not in the MEPS:
  - MEPS non-respondents
  - Deceased NHIS respondents
  - NHIS respondents that moved and cannot be located
  - Persons institutionalized after the NHIS interview

Control totals are noted in the link file documentation.

Linked MEPS-NHIS Estimation Issues

MEPS 2002 Linkage File Record Counts				
	Linked to NHIS		Not Linked to NHIS	Total
	2000 NHIS PUF (n=100,617)	2001 NHIS PUF (n=100,760)		
MEPS HC-062 Panel 6 persons (n=21,959)	19,351 (88.1%)		2,608 (11.9%)	21,959
MEPS HC-062 Panel 7 persons (n=17,206)		15,577 (90.5%)	1,629 (9.5%)	17,206
Total	34,928 (89.2%)		4,237 (10.8)	39,165

Linking to the Area Resource File and Other Secondary Data

- **Geographic codes (geo-codes) for merging the MEPS data to external data sources are available for use with approved AHRQ Data Center projects.**
- **Geo-codes are derived from MEPS sample person's household address.**
- **Geo-codes include the Census Block-Group, Census Tract, State and County, Designated Market Area, and Census Place.**

The AHRQ Data Center is a physical space at AHRQ in Rockville, Maryland, where researchers with approved projects can be allowed access to files not available for public use. These files may contain data that AHRQ has not fully edited, or may contain levels of detail that are not approved for public release. Researchers with approved projects are allowed to access only the information required to complete their projects, and microdata files are not released. Summary data (tables, regression equations) may be removed from the Data Center subject to review by AHRQ staff. Regression output tends to be approved quickly while summary data may take a few days.

Linking to the Area Resource File, and Other Secondary Data

- Latitude and longitude are available for creating distance measures.
- The 1998 and 1999 MEPS only have broader state and county geo-codes available.
- No direct access to variables by Data Center users.
- Contractor creates merged data sets.

Linking to the Area Resource File, and Other Secondary Data

- What types of data can be merged?
  - Data from the Area Resource File (ARF)
  - Data from the Census Bureau
  - Any other state, county, tract, etc., data.
    For example:
    - State Medicaid funding data
    - State and/or county poverty measures
    - EPA pollution data

Linking to the Area Resource File, and Other Secondary Data

- User supplied data must have data for all geographic units.
  - For example, a data user may provide a data set with Medicaid-related variables for each state.
  - In this case, there cannot be any missing values for any state, including the District of Columbia.

Area Resource File (ARF) Specifics

- **What is the ARF?**
  - **Database of compiled secondary data for each county in the US.**
  - **Contains over 6,000 variables.**
  - **Used for health service research, health policy analysis, and other geographically based activities.**
  - **Contains supply-side information at the state and local level; for example, level of HMO penetration, number of doctors, etc.**

Data sources for the ARF include NCHS detail mortality and natality records, AHA facilities, and AMA physician specialty data. All information contained on the file is derived from existing data sources.

HRSA -- Health Resources and Services Administration: http://www.hrsa.gov

Area Resource File (ARF) Specifics (cont'd)

- Updated each year.
- Codebook can be requested from the Data Center manager.
- The Web site for the ARF is http://www.arfsys.com/main.htm

Types of Projects That Utilize Geo-Codes

- Generally, Data Center projects that use geographic codes either
  - Examine the impact of policy or funding level changes to health care programs at the state level
  - Examine access to care issues, usually at the county level, typically using data from the Area Resource File

Events Files Linkage

- There are eight different event files:
  - Prescribed Medicines
  - Hospital Inpatient Stays
  - Outpatient Visits
  - Home Health
  - Emergency Room Visits
  - Office-Based Medical Provider Visits
  - Dental Visits
  - Other Medical Expenses

Events Files Linkage

- Each record represents a unique event and its attributes for a person as reported by a household respondent and includes
  - total expenditure
  - source of payment
  - up to 4 ICD-9 codes (except Home Health, Dental and Other Medical).
- Depending on the number of events reported, persons may be represented on the file
  - once
  - several times
  - not at all

For example, the following variables are from file HC-059G, MEPS 2001 OFFICE-BASED MEDICAL PROVIDER VISITS:

OBMD01X -- AMOUNT PAID, MEDICAID (IMPUTED)
OBMR01X -- AMOUNT PAID, MEDICARE (IMPUTED)
OBOF01X -- AMOUNT PAID, OTHER FEDERAL (IMPUTED)
OBOR01X -- AMOUNT PAID, OTHER PRIVATE (IMPUTED)
OBOT01X -- AMOUNT PAID, OTHER INSURANCE (IMPUTED)
OBOU01X -- AMOUNT PAID, OTHER PUBLIC (IMPUTED)
OBPV01X -- AMOUNT PAID, PRIVATE INSURANCE (IMPUTED)
OBSF01X -- AMOUNT PAID, FAMILY (IMPUTED)
OBSL01X -- AMOUNT PAID, STATE & LOCAL GOV (IMPUTED)
OBTC01X -- HHLD REPORTED TOTAL CHARGE (IMPUTED)
OBTR01X -- AMOUNT PAID, TRICARE (IMPUTED)
OBVA01X -- AMOUNT PAID, VETERANS (IMPUTED)
OBWC01X -- AMOUNT PAID, WORKERS COMP (IMPUTED)
OBXP01X -- SUM OF OBSF01X - OBOT01X (IMPUTED)

OBICD1X -- 3-DIGIT ICD-9-CM CONDITION CODE
OBICD2X -- 3-DIGIT ICD-9-CM CONDITION CODE
OBICD3X -- 3-DIGIT ICD-9-CM CONDITION CODE
OBICD4X -- 3-DIGIT ICD-9-CM CONDITION CODE

Events Files Linkage

- **Record Identifiers**
  - **DUPERSID identifies the person associated with the event.**
  - **EVNTIDX uniquely identifies each event (except for the Prescribed Medicines events).**
  - **LINKIDX uniquely identifies each Prescribed Medicines event.**
  - **FFEEIDX uniquely identifies a flat fee group. For example, pregnancy is typically covered in a flat fee arrangement.**
  - **Event files do not contain any demographic data. They must be linked to person-level files using DUPERSID to pull over information.**

If covered under a flat fee, all pregnancy-related events (the prenatal visit, the delivery, and the postpartum visits) would have the same value for FFEEIDX.

Events Files Linkage

- **Example: Using the 2001 Office-Based Medical Provider Visits File, create a person-level variable to identify persons 18 years old and older that had at least one office visit for a general checkup.**

General checkup adults 18 and older	Sample size	Percent	SE percent
Total	22965	100.0	0.0
No general checkup	13596	57.5	0.52
At least one general checkup	9369	42.5	0.52

Then later, cross that measure with health insurance status and calculate average total and family expenditure for the visit, by health insurance status.

File HC-059G, the 2001 Office-Based Medical Provider Visits file. The variable VSTCTGRY (question MV07 in the Medical Provider Visits questionnaire) indicates the category of care for the visit:

MV07
Please look at this card and tell me which category **best** describes the care (PERSON) received during the visit to (PROVIDER) on (VISIT DATE)?

GENERAL CHECKUP..... 1
DIAGNOSIS OR TREATMENT..... 2
EMERGENCY (E.G., ACCIDENT OR INJURY).....3
PSYCHOTHERAPY OR MENTAL HEALTH COUNSELING..... 4
FOLLOW-UP OR POST-OPERATIVE VISIT..... 5
IMMUNIZATIONS OR SHOTS..... 6
VISION EXAM..... 7
MATERNITY CARE (PRE/POSTNATAL)..... 8
WELL CHILD EXAM.....9
OTHER..... 91
REF..... -7
DK..... -8

Events Files Linkage

- General checkup office-based medical provider visit by health insurance status, adults 18 years old and older

General Checkup		Health Insurance Status		
		Total	Insured	Uninsured
Total	Sample Size	22965	19315	3650
	Col Percent	100.0	100.0	100.0
	SE Col Percent	0.0	0.0	0.0
No general checkup	Sample Size	13596	10576	3020
	Col Percent	57.5	53.9	82.1
	SE Col Percent	0.5	0.5	0.8
At least one general checkup	Sample Size	9369	8739	630
	Col Percent	42.5	46.1	17.9
	SE Col Percent	0.5	0.5	0.8

Health Insurance status comes from the full-year file.

Events Files Linkage

- **Average amount paid for an office-based medical provider visit for a general checkup, by insurance status, adults 18 years old and older**

Health insurance status	Sample size	Label	Mean
Insured	8739	Total paid	$271.70
		Paid by family	38.73
Uninsured	630	Total paid	$204.03
		Paid by family	83.54

Conditions File Linkage

- Each record represents a unique condition or procedure for a sample person, reported by a household respondent.
- Depending on the number of conditions they reported, persons may be represented on the file
  - once
  - several times
  - not at all

Conditions File Linkage

- Interviewer records verbatim text reported by the household respondent.
  - Open-ended questions
    - condition enumeration section
    - medical event sections
    - disability section
- Coded in post-processing by professional coders into ICD-9 codes.

Conditions File Linkage

- **Respondents may report having the same condition more than once.**
  - **Interviewer verifies that these are different occurrences of the condition.**
  - **Each unique episode of a condition is recorded only once.**
    - **Person may have more than one cold in a year.**
    - **Each cold has a separate record.**
  - **Chronic conditions should only be reported once. However, the file was not edited to ensure that no duplicates were recorded.**

For example, a chronic condition like diabetes should be reported only once for any given person. However, the condition of diabetes can be reported in several different places in the data collection process. For each subsequent report of diabetes, the respondent will be asked if this is the same diabetes as reported earlier. If the respondent states that this is a different diabetes, the subsequent report of diabetes will be treated as a different condition and given a separate record on the Conditions File.

Conditions File Linkage: Estimation Issues/Caveats

- **Analysts should not presume a high level of precision in condition data:**
  - **Inaccurate or vague reports of condition**
  - **Clustering of ICD-9 codes in NEC (not elsewhere classified)**
  - **One respondent provides information for the entire household**

Conditions File Linkage: Estimation Issues/Caveats

- For reasons of confidentiality, the ICD-9 codes on the PUF are three digits. Four and five digit ICD-9 codes are available for use with approved CFACT Data Center projects.
  - MEPS was not designed to make prevalence estimates for conditions.
- This file does not contain condition data collected in round 3/5 priority conditions section of the instrument.

Conditions File Linkage

- **Record Identifiers**
  - **DUPERSID identifies the person associated with the condition.**
  - **CONDN indicates the condition number (e.g., condition number 1,2,3, etc.) as it was reported during the interview for an individual respondent plus a control digit.**
  - **CONDIDX uniquely identifies each condition (i.e., each record on the file) and is the combination of DUPERSID and CONDN.**
  - **CONDRN indicates the round in which the condition was <u>first</u> reported.**

Conditions File Linkage: Identify Persons with Asthma

- **Example: Using the 2001 Medical Conditions file, create a person-level variable to identify persons that have at least one record for asthma.**

Asthma (ICD-9 = 493)	Sample size	Percent
Total	32122	100.0
Has asthma	1584	4.9
Does not have asthma	30538	95.1

The 2001 Medical Conditions file -- file HC-061

Linking Conditions to Events

- Event files only contain up to four conditions, in the order they were reported. To get all conditions, you must link to the conditions file.
- Events for Multiple Conditions
  - Events may be associated with more than one condition.
  - Example: One hospital stay for three conditions
    - Fractured hip, fractured shoulder, concussion
- Conditions for Multiple Events
  - Conditions may be associated with more than one event.
  - Example: Fractured hip has three events
    - Hospital stay, office-based medical provider visit, pharmacy visit

Linking Conditions to Events

- **Two Appendix Files**
  - **File 1 links each record on the condition file with one or more records on the event files.**
  - **File 2 links records on the Prescribed Medicines file with one or more records on the event files.**
- **File 1: Condition-Event Link File (CLNK) – use the CLNK file to link conditions to all associated event files.**

CLNK File

- **Record Identifiers on the CLNK file**
  - **DUPERSID identifies the person associated with each record.**
  - **EVNTIDX uniquely identifies each event for a person and corresponds to a unique record on one of the event files.**
  - **CONDIDX uniquely identifies each condition for a person and corresponds to a unique record on the Condition file and is the combination of DUPERSID and CONDN.**
  - **CLNKIDX uniquely identifies each record on the CLNK file and is the combination of CONDIDX + EVNTIDX.**
  - **EVENTYPE indicates the type of event record identified by EVNTIDX.**

MEPS H59IF1 CODEBOOK
2001 CONDITION-EVENT LINK FILE
DATE: March 17, 2004
NAME: EVENTYPE
DESCRIPTION: TYPE OF EVENT CONDITION IS LINKED TO

VALUE	UNWEIGHTED
1 MVIS	157,334
2 OPAT	17,345
3 EROM	6,866
4 STAZ	4,060
7 HVIS	7,981
8 PMED	121,013
TOTAL	314,599

CLNK File

The following records are for one randomly selected person from the 2001 conditions file, CLNK file, and associated events files:

Conditions file		CLNK file			Events file	
					PMED Events	
DUPERSID	CONDIDX	DUPERSID	CONDIDX	EVNTIDX	DUPERSID	LINKIDX /EVNTIDX
42690023	426900230010	42690023	426900230010	426900230023	42690023	426900230030
42690023	426900230026	42690023	426900230010	426900230030	42690023	426900230047
42690023	426900230032	42690023	426900230026	426900230047	42690023	426900230105
42690023	426900230048	42690023	426900230026	426900230061	42690023	426900230105
42690023	426900230054	42690023	426900230026	426900230078	42690023	426900230105
42690023	426900230060	42690023	426900230026	426900230105	42690023	426900230105
		42690023	426900230048	426900230061	42690023	426900230105
		42690023	426900230048	426900230078	42690023	426900230105
		42690023	426900230054	426900230085	42690023	426900230112
		42690023	426900230054	426900230092	42690023	426900230129
		42690023	426900230054	426900230112	Office-based Events	
		42690023	426900230054	426900230129	42690023	426900230023
					42690023	426900230061
					42690023	426900230078
					42690023	426900230085
					42690023	426900230092

RXLX File

- File 2: Prescribed Medicines-Event Link File (RXLK) – used to link each record on the Prescribed Medicines file with one or more records on the event files.
- Record Identifiers on the RXLK file:
  - DUPERSID identifies the person associated with each record.
  - EVNTIDX uniquely identifies each event for a person and corresponds to a unique record on one of the event files.

RXLX File

- **LINKIDX identifies the record(s) on the Prescribed Medicines file which link to an event record. There may be more than one record on the RXLK file for a specific LINKIDX value, and there may be more than one record on the Prescribed Medicines file for a specific LINKIDX value.**
- **RXLKIDX uniquely identifies each record on the RXLK file and is the combination of EVNTIDX + LINKIDX. There is just one record on this file for each value of RXLKIDX, i.e., each unique combination of EVNTIDX + LINKIDX.**
- **EVENTYPE indicates the type of event record identified by EVNTIDX.**

MEPS H59IF2 CODEBOOK
2001 PRESCRIBED MEDICINES - EVENT LINK FILE
DATE: March 17, 2004
NAME: EVENTYPE
DESCRIPTION: TYPE OF EVENT RX IS LINKED TO

VALUE	UNWEIGHTED
1 MVIS	47,737
2 OPAT	1,946
3 EROM	3,407
4 STAZ	3,355
5 DVIS	2,318
6 OMED	1,756
TOTAL	60,519

RXLX File

The following records are for one randomly selected person from the 2001 Prescribed Medicines file, RXLK file, and associated events files:

Prescribed Medicines file		RXLK file			Events file	
					Office-based Events	
DUPERSID	LINKIDX	DUPERSID	LINKIDX	EVNTIDX	DUPERSID	EVNTIDX
42690023	426900230030	42690023	426900230105	426900230061	42690023	426900230061
42690023	426900230047	42690023	426900230105	426900230078	42690023	426900230078
42690023	426900230105	42690023	426900230112	426900230085	42690023	426900230078
42690023	426900230105	42690023	426900230129	426900230092	42690023	426900230078
42690023	426900230105				42690023	426900230078
42690023	426900230105				42690023	426900230078
42690023	426900230105				42690023	426900230085
42690023	426900230105				42690023	426900230092
42690023	426900230112					
42690023	426900230129					

Linking Conditions to Events

- **Example: Identify persons with asthma and calculate the total expenditure for prescriptions, and amount paid by family, by insurance status.**

Average Prescription Expenditure per Person, for Persons with Asthma – Total and Paid by Family, by Health Insurance Status

Health Insurance Status	Sample size	Label	Mean
Any Private	798	Total Paid	431.10
		Paid by Family	139.14
Public Only	420	Total Paid	407.41
		Paid by Family	142.36
Uninsured	88	Total Paid	494.28
		Paid by Family	480.32

Linking to the Jobs File

- Each Jobs file covers one full year of data.
- One record for each job reported, in each round, for persons 16 years old and older.
- Many jobs have multiple records.
- Records contain details of job; i.e., wage rate, industry, occupation.
- Jobs information is collected using a dependent interview method.
- The information on the Jobs file is minimally edited.

In order to obtain complete information for a job, users must note the round in which the job is first reported. This is because MEPS collects complete Jobs information in the round in which a job is first reported.

In subsequent rounds, if job status remained the same, only a subset of the employment questions are asked, and many job-level variables on the subsequent round job records are coded as inapplicable (-1); the complete information is on the record for the job in the first round in which it was reported.

Linking to the Jobs File

- The Jobs file does not include any weights necessary to extrapolate this data to the U.S. population. To make person-level estimates, link to any of the other MEPS files and use the person-level weight on that file.
- Record Identifiers:
  - DUPERSID – Identifies the person associated with the job record.
  - JOBSN – Job number
  - JOBSIDX - The unique record identifier in the Jobs file is comprised of a person identifier (DUPERSID), a round identifier (RN), and a job number (JOBSN).

Linking to the Jobs File

- **Each job record contains variables related to the employment section of the MEPS household survey.**
- **For all persons age 16 and older, all jobs held within a round are recorded.**
- **The person-level file contains a limited number of variables related to a person's "current main job." The Jobs file contains complete information for all jobs.**

Persons who held more than one job at the round's interview date were asked to identify the main job. This job was classified as the "current main job" and all other simultaneously held jobs were classified as "miscellaneous." The MEPS also obtained some information on former jobs held in the reference period and, for those persons not currently working and having no job in the reference period, some information on the last job the person held. Additionally, for those persons age 55 or older who indicated that they retired from a job, the MEPS obtained some job-level data on the retirement job.

To get more detailed information about the CMJ, or information about "miscellaneous" jobs, you must link to the Jobs file.

Linking to the Jobs File

- **Depending on the number of jobs reported, persons may be represented on the Jobs file**
  - **once**
  - **several times**
  - **not at all**

Linking to the Jobs File

- **Example: Using the 2001 Jobs file, create a person-level variable for the number of jobs held by persons 18 years old and older using round 3/1 data.**

Number of Jobs	Sample size	Percent	SE percent
Total	18,788	100.0	0.0
0 Jobs	4,253	21.7	0.5
1 Job	13,238	70.7	0.5
2 Jobs	1,211	7.0	0.2
3+ Jobs	86	0.6	0.1

The 2001 Jobs file -- file HC-056